

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:20:32 ; Search time 17.21 Seconds

(without alignments)
123.847 Million cell updates/sec

Title: US-09-331-631a-24_COPY_29_94

Perfect score: 382
Sequence: 1 HDEDDRGGHSLQDCVQRC.....EQEEQGRGHWGEGERE 66

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwisProt_39:*

Pred. No. is the number of result's predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	133	34.8	573	1	GLBL_MAIZE
2	90	23.6	588	1	WCLB_GOSHI
3	85	22.3	639	1	GLCA_SOYBN
4	83.5	21.9	605	1	VCLA_GOSHI
5	76.5	20.0	338	1	LEGB_PEA
6	73	19.1	605	1	GLCA_SOYBN
7	72	18.8	852	1	SRCH_RABIT
8	70.5	18.5	185	1	12_MOUSE
9	70	18.3	1130	1	REPT_MOUSE
10	70	18.3	1224	1	PER_DROME
11	70	18.3	1549	1	TRHY_SHEEP
12	69.5	18.2	336	1	FILE_MOUSE
13	69.5	18.2	1208	1	PER_DROYA
14	69	18.1	661	1	PER_DROSE
15	68.5	17.9	654	1	CIR4_MOUSE
16	68.5	17.9	655	1	CIR4_RAT
17	68	17.8	654	1	CIR4_MOUSE
18	67.5	17.7	1394	1	CNGA_BOVIN
19	67	17.5	228	1	DH2_ORYSA
20	67	17.5	516	1	LEGB_GOSHI
21	67	17.5	653	1	CIR4_HUMAN
22	67	17.5	1121	1	DDX8_ARATH
23	66.5	17.4	656	1	PER_DROSI
24	66.5	17.4	676	1	PER_DROMA
25	66	17.3	407	1	LEGB_HSVSA
26	66	17.3	1407	1	TRHY_RABIT
27	65.5	17.1	412	1	TRHY_RABIT
28	65.5	17.1	1898	1	YNP1_CAEBL
29	65	17.0	355	1	SRP_MUSSP
30	65	17.0	524	1	SRP_SOYBN
31	65	17.0	888	1	YCB4_YEAST
32	65	17.0	1063	1	SPT5_YEAST
33	65	17.0	2248	1	CYAL_DROME

34	64.5	16.9	242	1	THYL_HUMAN	P20396	homo sapien
35	64.5	16.9	355	1	CPDI_DROME	P22058	drosophila
36	64	16.8	493	1	11S3_HELAN	P19084	heiliantus
37	64	16.8	543	1	P2CG_BOVIN	P79126	bos taurus
38	64	16.8	681	1	VE1_HPV10	P36720	human papil
39	64	16.8	797	1	VG48_HSVSA	001033	herpesvirus
40	63.5	16.6	708	1	GBF_DICDI	P36417	dictyostel
41	63.5	16.6	905	1	SNE5_YEAST	P18480	saccharomyc
42	63.5	16.6	1021	1	MANA_RHOMR	P49425	rhodothermu
43	63	16.5	699	1	SRCH_HUMAN	P23327	homo sapien
44	62	16.2	178	1	2SS2_BRANA	P01090	brassica na
45	62	16.2	258	1	2SS2_RICCO	P01089	ricinus com

ALIGNMENTS

RESULT	ID	GLBL_MAIZE	STANDARD:	PRT:	573 AA.
AC	P15590:				
DT	01-APR-1990 (Rel. 14, Created)				
DT	01-AUG-1990 (Rel. 15, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	GLOBALIN-1 S ALLELE PRECURSOR (GLBL-S) (7S-Like).				
GN	GLBL				
OS	Zea mays (Maize).				
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;				
OC	Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-CV. INBRED LINE VA26;				
RA	Belanger F.C., Kriz A.L.;				
RT	"Molecular characterization of the major maize embryo globulin encoded by the Glb1 gene."				
RT	[2]				
RL	Plant Physiol. 91:636-643(1989).				
RN	SEQUENCE OF 87-100.				
RP	MEDLINE=89374022; PubMed=2775172;				
RA	Kriz A.L.;				
RT	"Characterization of embryo globulins encoded by the maize Glb1 genes."				
RL	Biochem. Genet. 27:239-251(1989).				
CC	- I - PPM: THREE PROTEIN-PROCESSING STEPS OCCUR IN THE FORMATION OF THE MATURE PROTEIN FROM THE PRIMARY TRANSLATION PRODUCT.				
CC	- I - POLYMORPHISM: THE THREE MOST COMMONLY OCCURRING GLBL ALLELES HAVE THE DESIGNATION L, I, AND S FOR LARGE, INTERMEDIATE, AND SMALL PROTEINS, RESPECTIVELY.				
CC	- I - SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN, CONVICILIN, CONGLYCININ, ETC.).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL; M24845; AAA3467.1; -.				
DR	HSSP; P50477; ICAM.				
DR	MAITEDB; 30181; -.				
DR	INTERPRO: IPR001113; -.				
DR	PFAM; PF00546; Seedstore_7s; 1.				
FW	Seed storage protein; Signal.				
FT	SIGNAL	1	18		
FT	PROPEP	19	86		
FT	CHLIN	87	573		
FT	CARBOHYD	349	349		
SO	SEQUENCE	573 AA;	65029 MW;	525EDID0A062976 CXC64;	

Query Match 34.8%; Score 133; DB 1; Length 573;

[illegible]

```

DT 01-OCT-1969 (Rel. 12, Created)
DT 01-OCT-1969 (Rel. 12, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE BETA-CONGLYCININ, ALPHA CHAIN PRECURSOR.
GN CG-1.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
RN [1]
RX MEDLINE=86250867; PubMed=3013879;
RA Doyle J.J., Schuller M.A., Godelle W.D., Zenger V., Beachy R.N.,
RA Slightom J.L.;
RT "The glycosylated seed storage proteins of Glycine max and Phaseolus
RT vulgaris. Structural homologies of genes and proteins.";
RL J. Biol. Chem. 261:9228-9238(1986).
RN [2]
RP SEQUENCE OF 340-639 FROM N.A.
RX MEDLINE=83143289; PubMed=6897678;
RA Schuller M.A., Ladin B.P., Pollaco J.C., Freyer G., Beachy R.N.;
RT "Structural sequences are conserved in the genes coding for the
RT alpha, alpha' and beta-subunits of the soybean 7S seed storage
RT protein.";
RL Nucleic Acids Res. 10:8245-8261(1982).
CC -I- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED
CC DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A
CC CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.
CC -I- SUBUNIT: THE ALPHA-, ALPHA', AND BETA-SUBUNITS ASSOCIATE IN
CC VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.
CC -I- SUBCELLULAR LOCATION: CYTOLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC CC
CC -I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONLYCININ, ETC.).
CC CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL; M13759; AB01374.1; -
DR EMBL; J01290; -; NOT_ANNOTATED_CDS.
DR PIR; B24810; B24810.
DR HSSP; P02853; 2P8HL.
DR INTERPRO; IPR001113; -
DR PIRAM; PF005046; Seedstore_7s; 2.
KW seed storage protein; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 22
FT PROPEP 23 62
FT CHAIN 63 639 BETA-CONGLYCININ, ALPHA' CHAIN.
FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 551 551 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 543 543 P -> L (IN REF. 2).
FT CONFLICT 549 549 M -> V (IN REF. 2).
FT CONFLICT 608 608 S -> T (IN REF. 2).
FT SEQUENCE 639 AA; 74325 MW; 469BF24C79651E3F CRC64;

Query Match 22.3%; Score 85; DB 1; Length 639;
Best Local Similarity 24.3%; Pred. No. 0.081;
Matches 26; Conservative 9; Mismatches 20; Indels 52; Gaps 5;

OY 12 SLQGVQRCROEPRPY---SHARC-----VQECRDQ-----QGNRIIE 47
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 34 SHNKLCRSCNSEKSYRNQACNARCNLLKVEEECEGEGQIYRRPRQNRERQNGSKE 93
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 48 QEE-----EGGRCGMN-----GGEGEE 66
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 94 EDEGGRRRRRRRRPRQDNEENHOKKEENHWKKEENKGGSGEE 140
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

RESULT 4
VCLA_GOSHI
ID VCLA_GOSHI STANDARD; PRT; 605 AA.
AC P09799;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Borroto K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicilin) genes of cottonseed.";
RL Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19378; AAA33069.1; -
DR PIR: S06398; S06398.
DR HSSP: P50477; ICAX.
DR INTERPRO: IPR001113; -
DR PFAM: PF00546; Seedstore_7s; 1.
KW Seed storage protein; Signal.
FT SIGNAL 1 23
FT CHAIN 24 605 VICILIN GC72-A.
SQ SEQUENCE 605 AA; 71049 MW; C9DB9371C976953B CRC64;

Query Match 21.9%; Score 83.5; DB 1; Length 605;
Best Local Similarity 26.0%; Pred. No. 0.11;
Matches 20; Conservative 18; Mismatches 20; Indels 19; Gaps 4;

QY 1 HDDEDRGGSHLQOCVORC--ROERPRYSHARCVOECRDDQDQ-----HGRIHOE 49
ID GLCA_SOYBN STANDARD; PRT; 338 AA.
AC P13916;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE LEGUMIN B (FRAGMENT).
RN [1]
RP SEQUENCE FROM N.A.
RA LegB.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Pisum.
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Domoney C., Barker D., Casey R.;
RT "The complete deduced amino acid sequences of legumin beta-
RT polypeptides from different genetic loci in Pisum.";
RL Plant Mol. Biol. 7:467-474(1986).
CC -1- FUNCTION: THIS PROTEIN FOUND IN THE SEEDS OF MANY LEGUMINOUS &
CC NONLEGUMINOUS PLANTS IS THE SOURCE OF SULFUR-CONTAINING AMINO
CC ACIDS IN SEED MEALS.
CC -1- SUBUNIT: HEXAMER. EACH SUBUNIT IS COMPOSED OF AN ACIDIC AND A
CC BASIC CHAIN DERIVED FROM A SINGLE PRECURSOR AND LINKED BY A
CC DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE 11S SEED STORAGE PROTEINS (GLOBULINS)
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M16890; AAA33678.1; -
DR PIR: S04321; S04321.
DR INTERPRO: IPR000459; -
DR PFAM: PF00190; Seedstore_11s; 2.
DR PROSITE: PS00305; 11S_SEED_STORAGE; 1.
KW Seed storage protein; Multigene family.
FT NON_TER 1 1
FT CHAIN <1 161 ALPHA CHAIN (ACIDIC).
FT CHAIN 162 338 BETA CHAIN (BASIC).
FT DISULFID 168 168 INTERCHAIN (ALPHA-BETA) (POTENTIAL).
SQ SEQUENCE 338 AA; 38989 MW; 752CFC3D336B6A80 CRC64;

Query Match 20.0%; Score 76.5; DB 1; Length 338;
Best Local Similarity 22.7%; Pred. No. 0.33;
Matches 20; Conservative 17; Mismatches 26; Indels 25; Gaps 3;

QY 4 EDDRRGGSHLQOCVORCROERPRYSHARCVOECRDD-----QOQHGRHOEERGRG-- 55
ID GLCA_SOYBN STANDARD; PRT; 605 AA.
AC P13916;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE BETA-CONGLYCININ, ALPHA CHAIN PRECURSOR.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=COTYLEDON; Pubmed=2103438;
RX MEDLINE=91355860;
RA Sebastiani F.L., Farrell L.B., Schuler M.A., Beachy R.N.;
RT "Complete sequence of a cDNA of alpha subunit of soybean beta-
RT conglycinin.";
RL Plant Mol. Biol. 15:197-201(1990).
CC -1- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED
CC DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A
CC CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.
CC -1- SUBUNIT: THE ALPHA'-, ALPHA'', AND BETA-SUBUNITS ASSOCIATE IN
CC VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.
CC -1- SUBCELLULAR LOCATION: EMBRYO AXIS, AND COTYLEDONARY MEMBRANE-BOUND

```

```

QY 56 -----RGW-----HGEGREE 66
ID GLCA_SOYBN STANDARD; PRT; 605 AA.
AC P13916;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE BETA-CONGLYCININ, ALPHA CHAIN PRECURSOR.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=COTYLEDON; Pubmed=2103438;
RX MEDLINE=91355860;
RA Sebastiani F.L., Farrell L.B., Schuler M.A., Beachy R.N.;
RT "Complete sequence of a cDNA of alpha subunit of soybean beta-
RT conglycinin.";
RL Plant Mol. Biol. 15:197-201(1990).
CC -1- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED
CC DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A
CC CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.
CC -1- SUBUNIT: THE ALPHA'-, ALPHA'', AND BETA-SUBUNITS ASSOCIATE IN
CC VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.
CC -1- SUBCELLULAR LOCATION: EMBRYO AXIS, AND COTYLEDONARY MEMBRANE-BOUND

```

CC VACUOLAR PROTEIN BODIES.
CC -I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLICININ, ETC.).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X17698; CAA35691.1; -
CC PIR: S14681; FWSYBA.
CC HSSP: P50477; ICAM.
CC INTERPRO: IPR001113; -
CC PFAM: PF00546; Seedstore_7s; 1.
CC Seed storage protein; Signal; Glycoprotein; Multigene family.
CC SIGNAL 1 22
CC PROPEP 23 62
CC CHAIN 63 605 BETA-CONGLICININ, ALPHA CHAIN.
CC CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 517 517 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 605 AA; 70293 MW; CBEBA30506B8BC57 CRC64;
SO

Query Match 19.1%; Score 73; DB 1; Length 605;
Best Local Similarity 30.3%; Pred. No. 1.3;
Matches 20; Conservative 6; Mismatches 14; Indels 26; Gaps 3;

QY 15 OCVCORROERPRY-----SHARC-----VOEC-----RDQQQHGHEQ 48
DB 37 KCLQSCNSESDSYRNACARCNLKYKECEGEIIPRPPRPPHPPRQPGKEED 96
QY 49 EEEQGR 54
DB 97 EDEQPR 102

RESULT 7
SRCH_RABIT
ID SRCH_RABIT STANDARD; PRT; 852 AA.
AC P16230;
DT 01-APR-1990 (rel. 14, Created)
DT 01-APR-1990 (rel. 14, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE SARCOPLASMIC RETICULUM HISTIDINE-RICH CALCIUM-BINDING PROTEIN
DE PRECURSOR (HCP).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-SKELETAL MUSCLE;
RX MEDLINE-90036884; PubMed-2808365;
RA Hofmann S.L., Goldstein J.L., Orth K., Moormaw C.R., Staughter C.A.,
RA Brown M.S.;
RA "Molecular cloning of a histidine-rich Ca2+-binding protein of
RT sarcoplasmic reticulum that contains highly conserved repeated
RT elements.";
RL J. Biol. Chem. 264:18083-18090(1989).
CC -I- FUNCTION: HCP MAY PLAY A ROLE IN THE REGULATION OF CA(2+)
CC SEQUESTRATION OR RELEASE IN THE SR OF SKELETAL AND CARDIAC
CC MUSCLE.
CC -I- SUBCELLULAR LOCATION: SARCOPLASMIC RETICULUM LUMEN.
CC -I- SIMILARITY: STRONG TO HUMAN HRC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC or send an email to license@isb-sib.ch.
CC EMBL: J05080; AAA31279.1; -
CC PIR: A34373; A34373.
CC INTERPRO: IPR002134; -
CC PROSITE: PS00328; HCP; 10.
CC Calcium-binding; Signal; Repeat.
CC SIGNAL 1 27
CC CHAIN 28 852
CC MOD_RES 28 28
CC DOMAIN 59 100
CC REPEAT 59 79
CC REPEAT 80 100
CC DOMAIN 199 470 10 X TANDEM REPEATS, ACIDIC.
CC REPEAT 199 224
CC REPEAT 225 253
CC REPEAT 254 282
CC REPEAT 283 310
CC REPEAT 311 339
CC REPEAT 340 367
CC REPEAT 368 395
CC REPEAT 396 423
CC REPEAT 424 451
CC REPEAT 452 470
CC DOMAIN 471 585
CC DOMAIN 721 733
CC DOMAIN 780 826
CC SEQUENCE 852 AA; 96117 MW; A430C83BA94B30 CRC64;
SO

Query Match 18.8%; Score 72; DB 1; Length 852;
Best Local Similarity 27.8%; Pred. No. 2.3;
Matches 20; Conservative 11; Mismatches 21; Indels 20; Gaps 3;

QY 7 RRGHSLQOCVQRCORERPRYSARCVQRCRDQDQGR-----HEDEEGGRG----- 57
DB 663 KRGGSE-----EEEDGQTHHSHLEDEDEEGHGRSLQSDQDEBDRGESAKYOA 714
QY 58 ---WHGGEREE 66
DB 715 PLRHREEEEEE 726

RESULT 8
T2_MOUSE
ID T2_MOUSE STANDARD; PRT; 185 AA.
AC 006666;
DT 01-FEB-1995 (rel. 31, Created)
DT 01-FEB-1995 (rel. 31, Last sequence update)
DT 01-NOV-1997 (rel. 35, Last annotation update)
DE OCTAPEPTIDE-REPEAT PROTEIN T2.
GN SRST OR T2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-MACROPHAGE;
RX MEDLINE-93092084; PubMed-1458435;
RA di Carlo M., Montana G., Romancino D.P., Monteleone D.;
RA "A mouse repeat sequence conserved in eukaryotic genomes.";
RL J. Submicrosc. Cytol. Pathol. 24:467-472(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X67863; CAA48048.1; -

DR MGD: MGI:107677; Srst.
 FT Repeat.
 FT DOMAIN 2 159 20 X 8 AA APPROXIMATE TANDEM REPEATS.
 SO SEQUENCE 185 AA; 22805 MM; 8995BECAEC383971 CRC64;

Query Match 18.5%; Score 70.5; DB 1; Length 185;
 Best Local Similarity 35.5%; Pred. No. 0.76;
 Matches 22; Conservative 10; Mismatches 23; Indels 7; Gaps 3;

QY 6 DRGRGSHDQGVQRCRO-ERPRYSHARCVCQCRDDQOQHGHDEDEEGRGRGMHGEER 64
 Db 120 ERGRREARER--QGRGREARQGRBPA---EKGRGRQGRGVPAEAGRGKMKRQEA 173
 QY 65 EE 66
 Db 174 TE 175

RESULT 9
 ID REPT_MOUSE STANDARD; PRT: 1130 AA.
 AC P97347;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE REPERIN.
 GN RPTN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RC STRAIN-NMRI: TISSUE-SKIN;
 RX MEDLINE=97422611; PubMed=9268637;
 RA Kriegl P., Schuppler M., Koesters R., Mincheva A., Lichter P.,
 RA Marks F.;
 RT "Reperin (Rptn), a new member of the 'fused gene' subgroup within the
 RT S100 gene family encoding a murine epidermal differentiation
 RT protein.";
 RL Genomics 43:339-348(1997).
 CC -1- FUNCTION: NOVEL POTENTIAL PRECURSOR PROTEIN OF THE CORNIFIED CELL
 CC ENVELOPE.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
 CC -----
 CC "THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
 CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
 CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
 CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
 CC ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X99251; CAA67624.1; -
 DR HSSP: P02633; 3ICB.
 DR MGD: MGI:1099055; RPTN.
 DR INTERPRO: IPR001751; -
 DR INTERPRO: IPR002048; -
 DR PFAM: PF01023; S_100; 1.
 DR PFAM: PF00036; efhand; 1.
 DR PROSITE: PS00018; EF_HAND; 1.
 DR PROSITE: PS00303; S100_CABP; 1.
 KW Repeat; Calcium-binding.
 FT DOMAIN 1 91
 FT CA_BIND 22 33 S-100 LIKE.
 FT CA_BIND 62 73 SITE I (LOW AFFINITY) (POTENTIAL).
 SO SEQUENCE 1130 AA; 129884 MM; 7D97BDD25151918D CRC64;

Query Match 18.3%; Score 70; DB 1; Length 1130;
 Best Local Similarity 26.2%; Pred. No. 4.9;

Matches 17; Conservative 14; Mismatches 28; Indels 6; Gaps 2;

QY 1 HDDEDDRGGRGSHDQGVQRCROERPRYSHARCVCQCRDDQOQHGHDEDEEGRGRGMHGEER 60
 Db 991 HEEQSHQSDRQTHVDEQONQGRHROTH---EENHD--HQGRHHEDEHNHRQDHHQ 1044

QY 61 EGERE 65
 Db 1045 QGRQ 1049

RESULT 10
 ID PER_DROME STANDARD; PRT: 1224 AA.
 AC P07663; 017483; Q24446; Q24447; Q24448; Q24449;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PERIOD CIRCADIAN PROTEIN (CLOCK-6 PROTEIN) (CLK-6).
 GN PER.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RC STRAIN-OREGON-R;
 RX MEDLINE=87144607; PubMed=3102970;
 RA Cliri Y., Colot H.V., Jacquier A.C., Yu Q., Hall J.C., Baltimore D.,
 RA Rosbash M.;
 RT "A family of unusually spliced biologically active transcripts encoded
 RT by a Drosophila clock gene.";
 RL Nature 326:42-47(1987).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM PER-A/LONG).
 RC STRAIN-CANTON-S;
 RA Baylies M.K., Weiner L., Vossahl L.B., Saez L., Young M.W.;
 RT "Genetic, molecular and cellular studies of the period locus and its
 RT products in Drosophila melanogaster.";
 RL (in) Young M.W. (eds.);
 RL Molecular genetics of biological rhythms, pp.123-153, Marcel Dekker,
 RL New York (1993).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM PER-A/SHORT).
 RC STRAIN-CANTON-S;
 RX MEDLINE=86146900; PubMed=3081818;
 RA Jackson F.R., Bargiello T.A., Yun S.-H., Young M.W.;
 RT "Product of per locus of Drosophila shares homology with
 RT proteoglycans.";
 RL Nature 320:185-188(1986).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORMS PER-A/LONG; PER-A/SHORT; PER-D & PER-E).
 RA Murphy L., Harris D., Barrell B.G.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 499-1075 FROM N.A. (ISOFORM PER-A).
 RC STRAIN-OREGON-R;
 RX MEDLINE=86245055; PubMed=3087625;
 RA Reddy P., Jacquier A.C., Abovich N., Petersen G., Rosbash M.;
 RT "The period clock locus of D. melanogaster codes for a proteoglycan.";
 RL Cell 46:53-61(1986).
 RN [6]
 RP SEQUENCE OF 63-573 FROM N.A. (ISOFORM PER-A).
 RX MEDLINE=93170641; PubMed=8436278;
 RA Kilman R.M., Hey J.;
 RT "DNA sequence variation at the period locus within and among species
 RT of the Drosophila melanogaster complex.";
 RL Genetics 133:375-387(1993).
 RN [7]
 RP PHOSPHORYLATION.
 RX MEDLINE=94181572; PubMed=8134384;
 RA Edery I., Zwiabel L.J., Dembinska M.E., Rosbash M.;
 RT "Temporal phosphorylation of the Drosophila period protein.";


```

RT      "The gene for mouse epidermal filaggrin precursor. Its partial
RT      characterization, expression, and sequence of a repeating filaggrin
RT      unit.";
RL      J. Biol. Chem. 262:15643-15648(1987).
RN      [2]
RP      REVISONS.
RA      Rochnagel J.A.;
RL      Submitted (SEP-1988) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES
CC      DISULFIDE-BOND FORMATION AMONG THE INTERMEDIATE FILAMENTS DURING
CC      TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
CC      -1- PPM: FILAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
CC      HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES
CC      OF 248 AA, WHICH ARE NOT SEPARATED BY "LARGE LINKER". THE
CC      PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES. DURING TERMINAL
CC      DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on ways
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; J03458; AAA75559.1; -.
DR      PIR; A28444; A28444.
DR      MGD; MGI:95553; FIG.
KW      Phosphorylation; Polypeptide; Developmental protein.
FT      NON_TER
SO      SEQUENCE 336 AA; 35678 MW; 259F124D3AC0DB2D CRC64;

Query Match 18.2%; Score 69.5; DB 1; Length 336;
Best Local Similarity 27.7%; Pred. No. 1.7;
Matches 18; Conservative 12; Mismatches 26; Indels 9; Gaps 2;

OY      2 DDEDDRRCGHSIQCCVQRCQRCRPRYSARCVQECRDQDQGHGHEDEEGRG-RCGNG 60
DB      159 EESDSQHQHGHQHQHDEQHQHGHQHQH-----EHEQPSGHRQDQSSGRRGHQAHQ 210
OY      61 EGERE 65
DB      211 EGGRD 215

RESULT 13
PER_DROVA STANDARD; PRT; 1208 AA.
AC      024767; 026286;
DT      15-JUL-1999 (Rel. 38, Created)
DT      15-JUL-1999 (Rel. 38, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      PERIOD CIRCADIAN PROTEIN.
OS      Drosophila yakuba (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Thackeray J.R., Kyriacou C.P.;
RL      MEDLINE=91087242; Pubmed=1702156;
RL      "Molecular evolution in the Drosophila yakuba period locus.";
RL      J. Mol. Evol. 31:389-401(1990).
RN      [2]
RP      SEQUENCE OF 669-766 FROM N.A.
RA      MEDLINE=93140158; Pubmed=1487825;
RA      Peixoto A.A., Costa R., Wheeler D.A., Hall J.C., Kyriacou C.P.;
RT      "Evolution of the threonine-glycine repeat region of the period gene
RT      in the melanogaster species subgroup of Drosophila.";
J. Mol. Evol. 35:411-419(1992).
CC      -1- FUNCTION: ESSENTIAL FOR BIOLOGICAL CLOCK FUNCTIONS. DETERMINES THE

```

CC PERIOD LENGTH OF CIRCADIAN AND ULTRADIAN RHYTHMS; AN INCREASE IN
CC PER DOSAGE LEADS TO SHORTENED CIRCADIAN RHYTHMS AND A DECREASE IN
CC LEADS TO LENGTHENED CIRCADIAN RHYTHMS. ESSENTIAL FOR THE CIRCADIAN
CC RHYTHMICITY OF LOCOMOTOR ACTIVITY, ECLOSION BEHAVIOR, AND FOR THE
CC RHYTHMIC COMPONENT OF THE MALE COURTSHIP SONG THAT ORIGINATES IN
CC THE THORACIC NERVOUS SYSTEM. THE BIOLOGICAL CYCLE DEPENDS ON THE
CC RHYTHMIC FORMATION AND NUCLEAR LOCALIZATION OF THE TIM-PER
CC COMPLEX. LIGHT INDUCES THE DEGRADATION OF TIM, WHICH PROMOTES
CC ELIMINATION OF PER. NUCLEAR ACTIVITY OF THE HETERODIMER
CC COORDINATIVELY REGULATES PER AND TIM TRANSCRIPTION THROUGH A
CC NEGATIVE FEEDBACK LOOP. BEHAVES AS A NEGATIVE ELEMENT IN CIRCADIAN
CC TRANSCRIPTIONAL LOOP. DOES NOT APPEAR TO BIND DNA, SUGGESTING
CC INDIRECT TRANSCRIPTIONAL INHIBITION (BY SIMILARITY).
CC
CC -1- SUBUNIT: FORMS HETERODIMER WITH TIMELSS (TIM); THE COMPLEX THEN
CC TRANSLOCATES INTO THE NUCLEUS (BY SIMILARITY).
CC
CC -1- SUBCELLULAR LOCATION: NUCLEAR AT SPECIFIC PERIODS OF THE DAY.
CC FIRST ACCUMULATES IN THE PERINUCLEAR REGION ABOUT ONE HOUR BEFORE
CC TRANSLOCATION INTO THE NUCLEUS. INTERACTION WITH TIM IS REQUIRED
CC FOR NUCLEAR LOCALIZATION. (BY SIMILARITY).
CC
CC -1- DOMAIN: THE RUN OF GLY-THR IS IMPLICATED IN THE MAINTENANCE OF
CC CIRCADIAN PERIOD AT DIFFERENT TEMPERATURES. DELETION OF THE REPEAT
CC LEADS TO A SHORTENING OF THE COURTESHIP SONG CYCLE PERIOD, AND THUS
CC COULD BE IMPORTANT FOR DETERMINING FEATURES OF SPECIES-SPECIFIC
CC MATING BEHAVIOR. (BY SIMILARITY).
CC
CC -1- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY, PROBABLY BY THE
CC DOUBLE-TIME PROTEIN (DBT). PHOSPHORYLATION COULD BE IMPLICATED IN
CC THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER
CC PER-TIM (BY SIMILARITY).
CC
CC -1- SIMILARITY: TO OTHER G-T STRETCH CONTAINING PROTEINS.
CC
CC -1- SIMILARITY: CONTAINS A PAS (PER-ARMY-SIM) DIMERIZATION DOMAIN.
CC
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See <http://www.isdb.ch/announce/>
CC or send an email to license@isdb-sib.ch).
CC -----
CC DR EMBL; X61127; CAA43439.1; -.
CC DR EMBL; S53298; AAB25029.1; -.
CC DR FLYBASE; FBgn0013215; DyaA\per.
CC DR INTERPRO; IPR000104; -.
CC DR INTERPRO; IPR000104; -.
CC DR INTERPRO; IPR000104; -.
CC DR PFAM; PF00989; PAS; 2.
CC DR PRINTS; PRO0308; ANTIREFREZEI.
CC DR BIOLOGICAL_RHYTHMS; Repeat; Nuclear protein; Phosphorylation.
CC FT DOMAIN 56 79
CC FT REPEAT 231 299
CC FT REPEAT 382 449
CC FT REPEAT 463 503
CC FT DOMAIN 698 741
CC FT DOMAIN 74 79
CC FT DOMAIN 857 864
CC FT DOMAIN 889 893
CC FT DOMAIN 987 997
CC FT DOMAIN 1019 1025
CC FT DOMAIN 742 742
CC FT CONFLICT 751 751
CC FT CONFLICT 762 762
CC FT CONFLICT 762 762
CC SEQUENCE 1208 AA; 126820 MW; E534A35E473B8483 CRC64;

Query Match	18.2%	Score 69.5;	DB 1;	Length 1208;
Best Local Similarity	34.5%	Pred. No. 5.6;		
Matches 20;	Conservative	6;	Mismatches	19;
			Indels	13;
			Gaps	3;
QY	22	QERPRYSHARC	---VOECRD--DOOHGRHBOE-----EEOGRGRGWGEGEERE	66
		· : :	·	
Db	104	EERFRPSSGGGVBOQSCRELLPODQOHGEDHSEKATIEDQLQEGEDSGSSEAEARVE		161
RESULT	14			

DR EMBL: U03723: AAB60668.1; -.
 DR HSSP: P15385; 1ZTO.
 DR MGD: MGI:96661; KCNA4.
 DR INTERPRO: IPR000636; -.
 DR INTERPRO: IPR003091; -.
 DR PFAM: PF00520; Ion.trans; 1.
 DR PRINTS: PR00169; KCHANNEL.
 KW Ion channel; Transmembrane; Ion transport; Voltage-gated channel;
 Glycoprotein; Multigene family; Phosphorylation.
 FT TRANSMEM 309 327 SEGMENT S1 (POTENTIAL).
 FT TRANSMEM 372 393 SEGMENT S2 (POTENTIAL).
 FT TRANSMEM 405 425 SEGMENT S3 (POTENTIAL).
 FT TRANSMEM 444 462 SEGMENT S4 (POTENTIAL).
 FT TRANSMEM 479 498 SEGMENT S5 (POTENTIAL).
 FT TRANSMEM 540 562 SEGMENT S6 (POTENTIAL).
 FT DOMAIN 38 50 POLY-ALA.
 FT DOMAIN 62 65 POLY-HIS.
 FT DOMAIN 83 89 POLY-ARG.
 FT DOMAIN 123 137 POLY-GLU.
 FT DOMAIN 162 166 POLY-GLY.
 FT DOMAIN 434 437 POLY-GLN.
 FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT MOD_RES 600 600 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
 SQ SEQUENCE 654 AA; 73473 MW; 8693651A30BD36D4 CRC64;

Query Match 17.9%; Score 68.5; DB 1; Length 654;
 Best Local Similarity 27.3%; Pred. No. 4.1;
 Matches 15; Conservative 10; Mismatches 27; Indels 3; Gaps 1;
 QY 11 HSLQOCVQRCQERPRYSARCVQECRDDQOHGRHEQEEQGR--GRGMHSEG 62
 Db 96 HHROSFPHCSDLMPGSGSEKILRELSEEEDEDEEEEGRGFYFSEEDHDG 150

Search completed: March 1, 2001, 16:20:34
 Job time: 209 sec